Molecular biogeography of planktonic and benthic diatoms in the Yangtze River

Molecular Biogeography of Planktonic and Benthic Diatoms in the Yangtze River

Jiawen Wang[1†], Qingxiang Liu[1†], Xianfu Zhao[2], Alistair G. L. Borthwick[3], Yuxin Liu[1], Qian Chen[1,4], Jinren Ni[4,5*]

Author Affiliations:

[1]The Key Laboratory of Water and Sediment Sciences, Ministry of Education, College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, China

[2]Institute of Hydroecology, Ministry of Water Resources, Chinese Academy of Sciences, Wuhan 430079, China

[3]Institute of Infrastructure and Environment, School of Engineering, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JL, UK

[4]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining, 810016, China

[5]Beijing Innovation Center for Engineering Science and Advanced Technology, Peking University, Beijing, 100871, P. R. China

[†]Equal contributors

[*]Correspondence: Jinren Ni, College of Environmental Sciences and Engineering,

Peking University, Beijing, 100871, China. E-mail: jinrenni@pku.edu.cn

Abstract

Background: Diatoms are of great significance to primary productivity in oceans, yet little is known about their biogeographic distribution in oligotrophic rivers.

Results: With the help of high-throughput sequencing analysis of 279 samples from the Yangtze River, we provided the first integral biogeographic pattern of planktonic and benthic diatoms over a 6030 km continuum along the world's third largest river. Our study revealed spatial dissimilarity of diatoms under varying landforms, including plateau, mountain, foothill, basin, foothill-mountain, and plain regions, from the river source to the estuary. Environmental drivers of diatom communities were interpreted in terms of photosynthetically active radiation, temperature, channel slope and nutrients, and human interference. Typical benthic diatoms, such as *Pinnularia*, *Paralia* and *Aulacoseira*, experienced considerable reduction in relative abundance downstream of the Three Gorges Dam and the Xiluodu Dam, two of the world's largest dams.

Conclusions: Our study revealed that benthic diatoms are of particular significance in characterizing motile species in the riverine environment, thus providing insight into diatom biogeography and biogeochemical cycles in large river ecosystems.

Keywords: diatoms, biogeography, water, sediment, landform, environmental driver, human interference, Yangtze River

Background

Diatoms play a particularly important role in the biogeochemical cycle [1] of primary elements such as carbon, nitrogen, phosphorus, and silica, contributing about 20-25 % of global primary production [2]. Diatoms are ubiquitous and diverse species of single-celled, eukaryotic, photosynthetic microorganisms on earth [3], and are often the dominant primary producers in marine and freshwater ecosystems [4]. Therefore, diatoms in such ecosystems may be remarkably dissimilar either in phylogenetic composition or biogeographic distribution [5,6]. Freshwater bodies typically consist of lentic (particularly lakes and wetlands) and lotic waters (including streams and rivers), which are often dominated respectively by planktonic algae and benthic species [7].

Accurate identification of diatoms depends on the reliability of the analysis methods used. Morphological analysis requires extensive taxonomic expertise, and may exhibit shortcomings in characterizing specific diatoms in rivers [8]. With the development of High-Throughput Sequencing (HTS) technology, DNA metabarcoding has become a rapid, accurate and reliable method for diatom detection [9]. Various DNA barcoding studies have been successfully conducted, based mainly on different maker genes, including COI [10], ITS [11], and 18S rDNA [9,12]. Malviya et al. [13] were the first to estimate the diversity and distribution of planktonic diatoms based on the V9 region of eukaryotic 18S rDNA. Recently, the V4 region of 18S rDNA was proposed for diatom barcoding in a study of diatoms in river and deltaic

63 systems [9,12].

64 Comparing the numerous studies of diatoms and eutrophication in oceans [14,15]
65 and lakes [16,17] to date, it is clear that the present understanding of diatoms is
66 relatively poor for lotic and oligotrophic rivers [7]. In fact, previous reports on the
67 dynamics of riverine diatoms have mostly focused on tributaries, small rivers,
68 reaches, stations, and estuaries [18–20]. Many studies examined the diversity and
69 composition of planktonic [19,20] or benthic diatoms [18,21,22] based on
70 morphological identification. For example, Centis et al. [20] investigated planktonic
71 diatoms dominated by physical constraints at two stations of the River Adige, Italy.
72 Liu et al. [18] investigated the community structure of benthic diatoms in the Dong
73 River, one of the three main tributaries of the Pearl River, China. Although Kireta et
74 al. [23] observed that both planktonic and periphytic diatoms could be used as bio-
75 indicators of river conditions, little is known about the distinction between planktonic
76 and benthic diatoms regarding their spatiotemporal distributions.

77 Biogeography studies aim to reveal the spatial and temporal distribution of
78 biodiversity, and provide insight into the mechanisms that generate and sustain
79 diversity [24]. Spatial dispersal and environmental selection processes are regarded as
80 essential drivers for the biogeographical pattern of a given bacterial community [25].
81 The former promotes movement of species and their establishment at a new location,
82 whereas the latter alters the abundance and composition of species, according to the
83 ability to survive and reproduce under local environmental conditions. A similar

explanation has been proved to apply to the biogeographical pattern of planktonic or benthic diatom communities in small rivers using morphological analysis [26–28]. However, it remains unclear how the integrated spatiotemporal distributions of planktonic and benthic diatom communities are shaped by spatial dispersal and environmental selection processes in large rivers subject to complex natural and anthropogenic impacts.

To close the above gap, we implemented large-scaled synchronous monitoring of diatom communities at 62 hydrologic stations over a 6030 km continuum of the Yangtze River in China. Consequently, we successfully provided the first molecular biogeographic pattern of both planktonic and benthic diatoms in this, the largest river in Asia (Fig.1). Meanwhile, environmental drivers of diatom communities were interpreted in terms of photosynthetic active radiation, temperature, channel slope, and nutrient conditions under varying landforms.

Results

Our study generated 8,602,620 V4 18S rDNA reads, in total, from 279 samples. All sequencing reads were classified into 3947 operational taxonomic units (OTUs) at a 97% similarity threshold, with 3144 OTUs well matching 454 diatom species in our reference database. Rarefaction curves (Additional file 1: Figure S1) together with high values of Good's coverage ranging from 0.9854 to 0.9992 illustrated that OTUs obtained by the current sequencing depth gave a reasonable representation of the diatom communities. The phylogeny tree, constructed by representative OTUs

(accounting for > 90% sequence in all samples) and reference sequences (Additional file 1: Figure S2), further confirmed the accuracy of taxonomic assignment.

Molecular insight into the diatom communities

Molecular barcoding based on high-throughput sequencing (HTS) provided a detailed diatom directory for the whole Yangtze River at different taxonomy levels, i.e., 4 classes, 37 orders, 60 families, and 152 genera.

HTS is of particular use in detecting nano-sized diatoms (2-20 μm) with several HTS studies for the Yangtze River confirming the presence of *Fragilaria perminuta*, *Achnanthidium minutissimum*, *Achnanthidium saprophilum*, *Amphora pediculus*, *Fistulifera saprophila*, *Mayamaea permitis*, *Sellaphora seminulum*, *Encyonema minutum*, *Fragilaria famelica*, *Fragilaria rumpens*, *Gomphonema pumilum*, *Staurosirella pinnata*, *Planothidium frequentissimum*, *Craticula buderi*, and *Craticula molestiformis*.

Molecular characterization of riverine diatoms revealed that specific diatoms had a statistically significant response to concentration of dissolved carbon dioxide (pCO_2, ppm) in the Yangtze River (Additional file 1: Figure S3). Planktonic diatoms *Asterionella formosa*, *Diatoma vulgare*, *Lindavia viaradiosa*, *Gomphonema pumilum*, and *Thalassiosira nordenskioeldii* were strongly associated with dissolved carbon dioxide (Spearman $r > 0.3$, $p < 0.05$), whereas benthic diatoms *Asterionella formosa*, *Encyonema prostratum*, *Eucocconeis laevis*, *Fistulifera saprophila*, and *Nitzschia sigmoidea* were highly correlated with dissolved carbon dioxide.

Moreover, fifteen genera (*Cyclotella*, *Achnanthes*, *Asterionella*, *Bacillaria*, *Cocconeis*, *Cymbella*, *Diatoma*, *Fragilaria*, *Gomphonema*, *Gyrosigma*, *Melosira*, *Navicula*, *Nitzschia*, *Rhizosolenia*, *Synedra*) were simultaneously detected in water-autumn samples by HTS and traditional morphological studies. Relative abundances of *Melosira* (19.90%), *Achnanthes* (18.55%), *Synedra* (11.84%), and *Navicula* (4.27%) identified by the morphological study were higher than those by molecular analysis (i.e., 15.12%, < 0.0001%, 0.01%, and 0.65%, respectively). *Cyclotella* dominated the riverine diatoms, as confirmed by both light microscopy (33.90% of total diatoms) and HTS (30.23% of total reads). Additional genera of considerable abundance were discovered in the Yangtze River solely by means of the HTS study, including the genus *Skeletonema* (relative abundance of 10.30%), *Paralia* (8.24%), and *Pinnularia* (4.98%).

Alpha and beta diversity of diatom communities

Six types of environmental samples were taken along the Yangtze River, including water and sediment samples from the river source region (i.e. water-plateau (12 samples) and sediment-plateau (12 samples)) and those from the mainstream in the non-plateau area (i.e. water-spring (38 samples), water-autumn (46 samples), sediment-spring (87 samples), and sediment-autumn (84 samples)). Planktonic diatoms exhibited the highest alpha-diversity (Chao1 and Shannon indices) and benthic diatoms the lowest richness (Chao1) in the plateau (Additional file 1: Figure S4). In the non-plateau area, no significant differences were observed in the alpha

147 richness and diversity of diatom communities in the four sample types.

148 Non-metric multidimensional scaling (NMDS) analysis of the compositional
149 dissimilarities between diatom communities of all samples demonstrated not only a
150 clear spatial differentiation in diatoms between the plateau and the main body of the
151 Yangtze, but also a division between planktonic and benthic groups (Additional file 1:
152 Figure S5). Seasonal difference in planktonic diatoms is found much more significant
153 than in benthic diatoms, as further confirmed by an Analysis of Similarity (ANOSIM)
154 test (Additional file 1: Figure S6). Moreover, one-way Analysis of Variance (one-way
155 ANOVA) indicated that more planktonic diatoms (42.75±13.98 % relative abundance,
156 primarily belonging to *Cyclotella*, *Stephanodiscus* and *Skeletonema*) than benthic
157 diatoms (16.58±5.06 % relative abundance, primarily belonging to *Pinnularia* and
158 *Stephanodiscus*) exhibited significant seasonal sensitivity (Additional file 1: Figure
159 S7).

160 **Biogeographic patterns of diatom communities**

161 Indicator species analysis revealed differences in diatom taxonomic composition
162 among the six types of environmental samples. The number of indicator diatom
163 species in the river ranged from 6 (sediment-spring) to 41 (water-plateau) (Additional
164 file 1: Table S1). Diatom communities in the plateau region were quite different from
165 those in the main body of the Yangtze River, as evidenced by the higher percentage of
166 top indicator species in water-plateau and sediment-plateau samples (Additional file
167 1: Figure S8). The average relative abundance of indicator species in the source area

168 exceeded 40%, and planktonic indicator species contributed more reads than benthic
169 indicator species. Furthermore, a number of indicator species belonging to
170 *Tabellariales* and *Hemiaulales* occurred in water-plateau and sediment-plateau
171 samples, respectively.
172 Species composition and spatial distribution of planktonic and benthic diatoms at
173 different taxonomy levels along the whole Yangtze River were then studied. At class
174 level, *Coscinodiscophyceae* dominated planktonic diatoms whereas *Bacillariophyceae*
175 prevailed in the mainstream sediment. In the plateau region, diatom communities
176 displayed similar structures dominated by *Bacillariophyceae* in both water and
177 sediment. In the lower reach, the greater abundance of planktonic *Bacillariophyceae*
178 demonstrated a seasonal difference (Additional file 1: Figure S9).
179 At genus level, the relative abundance contributed by dominant genera (top 20)
180 ranged from 55.6 to 83.6 % (Fig. 2**a**). Dominant benthic genera, *Pinnularia*,
181 *Cyclotella*, and *Navicula*, represented 14.2, 13.4, and 13.1 % of total sequences in
182 sediment-autumn samples. Benthic *Navicula* was adapted to wide ranges of
183 temperature and elevation, as evident by its dominance in sediment-plateau (17.1 %)
184 and sediment-spring (13.2 %) samples. *Cymbella* was prevalent in water-plateau
185 (17.7 %) and sediment-plateau (9.27 %) samples.
186 At species level, diatom composition in terms of ecological guilds showed spatial
187 dissimilarity in water and sediment (Additional file 1: Figure S10). Diatoms were
188 divided into four ecological guilds according to their biological traits, including low-

189 profile, high-profile, motile and planktic guilds in terms of different responses to
190 nutrients and dynamic disturbances [29–31] (Methods: Ecological guilds
191 classification). Benthic diatoms in the motile guild prevailed at most stations along
192 the whole river, whereas those in high-profile and planktic guilds dominated upstream
193 and downstream reaches, respectively. In addition, planktonic diatoms in the planktic
194 guild were predominant at most stations along the Yangtze River.

195 Over the 6030 km continuum from river source to mouth, landform type plays a
196 significant role in spatial differentiation of both planktonic and benthic diatom
197 communities. Referring to a previous study on landform types in the Yangtze [32],
198 planktonic diatoms represented by *Cymbella*, *Asterionella, Stephanodiscus*, *Melosira*,
199 *Cyclotella* and *Conticribra* and benthic diatoms represented by *Cymbella*, *Navicula*,
200 *Melosira*, *Conticribra, Cyclotella* and *Surirella* were respectively abundant in plateau,
201 mountain, foothill, basin, foothill-mountain, and plain regions (Fig. 2**b**).

202 **Natural and anthropogenic effects on diatom biogeography**

203 Natural effects were estimated, including spatial dispersal and environmental
204 selection processes for the biogeographical pattern of diatom communities.
205 Significant distance-decay in diatom similarity was observed according to
206 geographical distance (Additional file 1: Figure S11), with a greater gradient for water
207 (slope = −0.042) than for sediment (slope = −0.038) using least squares linear
208 regression. The Partial Mantel test demonstrated that both geographical and
209 environmental distances played important roles in constraining diatom composition

and distribution (Additional file 1: Table S2). Variation partitioning of diatom composition showed that a greater percentage (14.6-21.2%) could be explained by a purely environmental component than that (3.4-6.0%) of the total variation by a purely spatial component (Additional file 1: Figure S12), and a minor portion (0.4-5.4%) explained by spatial structured environmental heterogeneity, leaving the majority of the total variation (68.7-79.0%) inexplicable. Although environmental differentiation seems more important than spatial dispersion in shaping a diatom community, neither can fully explain the total variation in diatom composition. Common environmental parameters, including photosynthetically active radiation, temperature, channel slope, and nutrients, affect diatom communities undergoing spatial dispersal.

Photosynthetically active radiation (PAR, 400-700 nm), important as solar energy, is utilized by diatoms to synthesize biomass through photosynthesis [33]. Regardless of weak changes in PAR between spring and autumn in the Yangtze River basin [34], the spatial composition and distribution of both planktonic and benthic diatom communities varied in four PAR regions (Fig. 3**a-b**, Methods: Photosynthetically Active Radiation (PAR) divisions). LefSe analysis revealed that both planktonic and benthic diatom species preferred living under different conditions of light intensity. For example, the *Caloneis*, *Cymbella*, *Fistulifera* and *Fragilaria* genera preferred very-high PAR zones, the *Papiliocellulus* genus favored medium PAR regions, and *Conticribra* and *Cyclotella* lived as biomarkers in low PAR habitats. Planktonic

231 *Cymatopleura* and *Navicula,* and benthic *Asterionella*, *Biddulphia*, *Diatoma* and
232 *Encyonema* genera were adapted to high PAR conditions. Moreover, water
233 temperature is a key environmental factor in structuring diatom community
234 assemblages through its influence on diatom size and growth rate [35] in the Yangtze
235 River (Additional file 1: Table S3). Although the richness of planktonic diatoms
236 seems to fluctuate with PAR, the richness of benthic diatoms tends to rise with
237 increasing temperature (Fig. 3**c**).
238 Stream power, often simply characterized by the river channel slope or the product
239 of channel slope and flow discharge (except in plateau regions) [36], is another
240 important factor altering the spatial distribution of diatoms (data sourced from Chen et
241 al. 2001 [37]). During the wet season (autumn), the higher flow discharge weakens
242 the correlation between planktonic community similarity and channel slope, although
243 a stronger correlation between benthic community similarity and channel slope is
244 maintained due to higher mobility of the streambed (Fig. 4**a-b**). In general, the
245 varying channel slope along the Yangtze River could be simplified into three stages:
246 steep slope in mountainous reaches (station 1~2); moderate slope in upper reaches
247 (station 3~14); and mild slope in middle-lower reaches (station 15~24). In view of
248 their relative abundance, planktonic diatoms are characterised by *Psammothidium*,
249 *Nitzschia* and *Cymbella* for steep-slope environments, *Papiliocellulus* for moderate
250 slope, and *Mayamaea*; *Pinnularia* and *Surirella* for mild slope environments. Benthic
251 diatoms are represented by *Cocconeis*, *Entomoneis* and *Melosira* for steep slope

252 environments, *Fallacia*, *Psammothidium* and *Skeletonema* for moderate slopes, and
253 *Actinocyclus*, *Aulacoseira* and *Conticribra* for mild slopes (Additional file 1: Figure
254 S13). Furthermore, slope effects on diatoms may be identified in terms of ecological
255 guilds. Regardless of the diatoms in an unspecified ecological guild, planktonic
256 diatoms are dominant in the planktic guild. Interestingly, species in motile guilds
257 constitute the main component of benthic diatoms in the whole lotic river (Fig. 4**c**).
258 Nutrient conditions, represented by the ratio of total nitrogen to total phosphorus
259 (TN:TP), are of great significance to diatom growth. Noting that Redfield [38]
260 proposed that N:P = 16:1 provides an optimal criterion for algal production in the
261 oceans, we then investigated the alpha-diversity of diatom community response to
262 TN:TP in the Yangtze River. The alpha-diversity of planktonic diatoms could be partly
263 interpreted in terms of annual-averaged dissolved TN:TP (data range from 2005 to
264 2014) in water-spring (Adj $R^2 = 0.54$) and water-autumn (Adj $R^2 = 0.41$) samples
265 (Additional file 1: Figure S14**a-b**); however, benthic diatoms demonstrated an even
266 weaker response to TN:TP (monitored data) (Additional file 1: Figure S14**c-d**). For
267 relatively oligotrophic rivers like the Yangtze River, it appears that TN is more
268 important than TP as limiting nutrients to planktonic and benthic community
269 variability (Additional file 1: Table S3). Furthermore, the measured TN:TP provided a
270 better explanation for community variations in benthic rather than planktonic diatoms,
271 which could be attributed to the difference in utilization of phosphorus forms between
272 benthic and planktonic diatoms (Additional file 1: Table S3).

Nevertheless, the aforementioned environmental factors are subject to change due to human interference. For the Yangtze River, one of the most prominent impacts on its ecology arises from the construction and operation of large dams. In the present study, a one-way ANOVA analysis revealed that certain OTUs exhibited significant variations in distribution immediately upstream and downstream of the Three Gorges Dam ($p < 0.01$) and Xiluodu Dam ($p < 0.05$). Moreover, sudden drops in relative abundance of OTUs belonging to specific benthic species (such as *Pinnularia*, *Paralia* and *Aulacoseira*) occurred downstream of the dams (Additional file 1: Figure S15).

Discussion

The molecular method provided a comprehensive explanation of the molecular biogeographical pattern of diatom communities in a large river, the Yangtze. This is for three reasons. First, plentiful information on species composition of diatom communities is provided by the HTS method (which benefits from a continuously updated NCBI database). Second, HTS characterizes diatoms based on genetic information taken from the V4 region of 18S rDNA, enabling evolutionary relationships to be distinguished, and then elucidated with confidence. Third, the HTS method greatly facilitates the identification of small-sized genera that otherwise might be either regarded as unclassified species or overlooked due to their minute size and difficulty of detection. Species-specific diatoms of different cell size make different contributions to primary productivity and bloom biomass [39], whereas

294 nanoplanktonic diatoms may play a vital role in carbon export [40].

295 This study offers insights into specific planktonic and benthic diatoms, which are of
296 great importance to the carbon cycle. Here, the diatoms exhibited a significant
297 positive correlation with dissolved carbon dioxide [41]. In the river, diatoms take up
298 dissolved inorganic carbon and convert it into dissolved or particulate organic matter
299 (primary production), and then reduce carbon dioxide in the surface layer.
300 Conversely, re-mineralization of organic matter by bacteria and zooplankton can
301 increase carbon dioxide. Therefore, diatoms could maintain the equilibrium between
302 carbon dioxide in the surface layer with that in the atmosphere for the river system
303 [42].

304 The HTS method has revealed the existence of certain special species in the
305 Yangtze River. A poorly known freshwater species *Skeletonema potamos*, which has
306 high growth potential when silica is not limiting [43], was found to be a dominant
307 species of genus *Skeletonema* in the Yangtze. Although *Paralia* has previously been
308 reported to be abundant in marine, coastal, and river estuary areas [44,45], it has been
309 found in considerable abundance at several stations in the mid-lower Yangtze; this
310 may be attributed to saltwater intrusion aggravated by operation of the Three Gorges
311 Dam during the autumn [46]. *Pinnularia* has been shown to adapt to different habitats
312 with varying pH and nutrients, including moist soils, sediments, springs, and oceans
313 [47]. In the river system, a large number of *Pinnularia* species (such as *Pinnularia*
314 *viridiformis*, *P. subanglica*, *P. brebissonii*, *P. biceps*, etc.) were also detected, living in

315 weakly alkaline water and sediment environments.

316 Taxonomic compositions of riverine diatoms in the Yangtze River are markedly
317 different from those in lakes [48] and oceans [13] because of the distinct differences
318 in salinity and hydraulic conditions. For example, *Actinocyclus spp.*, *Aulacoseira spp.*,
319 *Cyclotella spp.*, *Fragilaria spp.*, and *Synedra spp.* predominated in diatom
320 communities in Lake Kasumigaura, Japan [48]. Moreover, *Chaetocero*s, followed by
321 *Fragilariopsis*, *Thalassiosira*, and *Corethron* were reported to be the most abundant
322 genera in the global oceans [13]. In marine ecosystems, diatoms are abundant in
323 nutrient-rich coastal zones particularly at high latitudes. In riverine ecosystems, the
324 spatial variation of most species can be explained in terms of landform type. For
325 example, diatoms of glacier origin from the Qinghai-Tibetan plateau were
326 significantly different to those in the main body of the Yangtze River (Fig. 2,
327 Additional file 1: Figure S8). In order of abundance, the most common riverine
328 diatoms in water and sediment of the Yangtze River were *Cyclotella*, *Navicula*,
329 *Pinnularia*, *Stephanodiscus,* and *Cymbella* genera.

330 With the help of synchronous water and sediment sampling along the Yangtze
331 River, it is possible to compare the biogeography of planktonic and benthic diatoms.
332 On the one hand, although the diversity of planktonic and benthic diatoms was very
333 similar in non-plateau areas, their community composition was remarkably different.
334 Planktonic diatoms were dominated by *Cyclotella* (average relative abundance:
335 17.66%), *Stephanodiscus* (12.81%), *Pinnularia* (7.53%), *Paralia* (7.34%), and

336 *Skeletonema* (4.90%); whereas benthic diatoms were dominated by *Navicula*
337 (13.12%), *Cyclotella* (10.33%), *Pinnularia* (10.12%), *Surirella* (7.10%) and
338 *Stephanodiscus* (6.53%). On the other hand, seasonal differences were more evident
339 in planktonic diatoms than benthic diatoms. A one-way ANOVA analysis confirmed
340 that the water temperature of the Yangtze River was significantly different between
341 spring and autumn ($p < 0.01$), with the water temperature (average 21 °C) in autumn
342 being more conducive than that (average 11 °C) in spring for planktonic diatom
343 growth [49]. Moreover, seasonal fluctuations in water discharge appear to affect the
344 community structure of planktonic diatoms, owing to the introduction of diatom
345 species from the upstream freshwater source and to different hydrologic processes in
346 spring and autumn [50]. Meanwhile, the weak seasonal difference of diatom
347 communities in sediment may be ascribed to the preponderance of benthic diatoms
348 (Additional file 1: Figure S6) that weakly responded to seasonal changes and reached
349 a state of relative equilibrium through long-term sediment erosion and deposition
350 processes [32].
351 To study the spatial distribution of planktonic and benthic diatoms, the Yangtze
352 River can be divided into six reaches according to the surrounding landform type [32].
353 In both water and sediment samples, differentially abundant diatoms were detected for
354 the six landform types across the sample site. In the Yangtze basin, landform changes
355 are associated with variations along the river in local soil type, soil nutrients, light,
356 altitude and temperature. The plateau reach is located at the Qinghai-Tibet Plateau

357 which has the lowest temperature (<11 degree centigrade) and the highest altitude
358 (>3500m) and light resource (PAR >32 $mol^{-2}m^{-1}d$). The basin reach is in Sichuan,
359 which receives the lowest light resource. The plain reach near the river estuary is
360 located in the Middle-Lower Yangtze Plain at the lowest altitude (< 10m) and the
361 highest temperature and nutrient levels (nitrogen and phosphorus). Therefore,
362 landform type, which determines the essential light, temperature, nutrients, and other
363 environmental factors for diatom growth, had a significant impact on the spatial
364 distribution of diatom community.
365 The environmental-based selection process driving the biogeographic pattern of
366 diatoms is influenced by PAR, temperature, channel slope, and nutrient level. Benthic
367 diatoms are typical consequences of riverine environmental selection in the Yangtze
368 River. We now consider the influence factors in turn. First, it has been reported that
369 sufficient PAR drives the growth and production of diatoms [51], but excess PAR can
370 affect various cellular processes and reduce the growth or viability of diatoms [52]. In
371 the present study, specific diatoms were identified as having adapted to different
372 levels of light intensity (Fig. 3), indicating that diatoms possess diverse light-
373 regulatory mechanisms and adaptive responses [53]. Although light and temperature
374 are the essential resource for diatom growth, planktonic and benthic diatoms exhibit
375 different preferences for PAR and water temperature. In the surface oceans,
376 planktonic diatoms have been shown to be replaced by small phytoplankton, causing
377 decreased primary production and carbon export, due to global warming [54]. In the

Yangtze River however, benthic diatoms are better biological indicators of temperature change. Second, as a primary driver of stream power that shapes the spatial distribution of diatoms, the channel slope not only affects competitive and succession processes among species but also alters nutritional utilization strategies and hence the production and growth of diatoms [55], leading to diatoms adapting to different flow conditions [56]. The different spatial distribution of ecological guilds suggests that benthic diatoms are appropriate markers of environmental heterogeneity along the river; however, the nutrient effect is not always significant on the dominant species in the motile guild under nutrient-oligotrophic conditions (Fig. 4). Third, the nutrient level characterized by annual-averaged dissolved TN:TP only partly explains the alpha-diversity of planktonic diatoms, whereas the monitored TN:TP provides a slightly better explanation for community variations of benthic diatoms. One reason could be related to species-specific responses to nitrogen and phosphorus in the production and growth of diatoms (Additional file 1: Figure S14 and Table S3). For example, *Nitzschia palea* [57] is phosphorus limited, but *Chaetoceros calcitrans* [58] is efficient in nitrogen assimilation. Another reason might be that algae utilize nitrogen and phosphorus in different forms; for example, ammonia is preferred to nitrate [59] and inorganic phosphate is preferred to organic phosphorus [60]. In the Yangtze River, the weak relationship between alpha-diversity of benthic diatoms and TN:TP reflects the lower influence of nutrients in this oligotrophic river.

Damming in rivers is a typical anthropogenic perturbation which could profoundly

399 modify material fluxes and biogeochemical cycles of downstream [61,62]. Kunz et al
400 [61] reported that sediment, carbon, nitrogen, and phosphorus were trapped by the
401 reservoir immediately upstream of the Itezhi-Tezhi Dam, increasing the N:P ratio
402 downstream of the dam. High flow downstream of dams has caused severe erosion of
403 the riverbed and led to coarsening of bed materials [62]. Changes in water level also
404 affect the light intensity and temperature to benthic diatoms. In short, a large dam
405 disrupts the hydraulic gradient, nutrient conditions, light availability, and temperature
406 in rivers, resulting in local changes to the environment in which benthic diatoms
407 thrive.

408 Interaction between planktonic and benthic diatoms is a major concern for large
409 river ecosystems. In the Yangtze river, such interactions have several consequences.
410 First, the community composition of planktonic community was significantly
411 correlated with that of benthic diatoms in paired water and sediment samples (in
412 spring: Spearman $r = 0.3556$, $p = 0.001$; in autumn: Spearman $r = 0.1902$, $p = 0.006$).
413 Typical benthic diatoms (e.g., *Nitzschia* and *Navicula*) were found in high abundance
414 in the water column. The local interactions could cause benthic and planktonic
415 habitats to become coupled through migration of algal cells, meaning that
416 phytoplankton can be derived from benthic diatoms, and sinking planktonic algae can
417 become benthic algae [63]. Second, the richness of planktonic diatoms appears to
418 fluctuate with PAR, whereas the richness of benthic diatoms tends to change with
419 temperature. Nutrient level (TN:TP) had different effects on the diversity and

variation of planktonic and benthic diatoms. These phenomena further explain the local interactions in terms of light, temperature and nutrients competition [64]. Third, given the “River continuum concept” [65] and “Continuous discontinuity concept” [66], the dominance of benthic or planktonic algae changes with the natural riverine gradient (e.g. channel slope), and interactions are invariably interrupted by anthropogenic disturbances such as dams, reservoirs, and nutrient-rich discharges. Other factors affecting interactions, such as river velocity, turbulent diffusion, algal sinking and grazing remain to be further investigated [67].

Conclusions

This study provided the first molecular biogeographic patterns of both planktonic and benthic diatoms over a continuum of 6030 km in the Yangtze River. Significant seasonal differences in planktonic diatom communities were observed over the whole length of the mainstream. Diatoms in water and sediment exhibited differential abundance according to landform type, such as plateau, mountain, foothill, basin, foothill-mountain, and plain regions, along the river. Environmental selection overcame spatial dispersal in controlling the community structure of planktonic and benthic diatoms. Typical environmental drivers were therefore interpreted in terms of photosynthetically active radiation, hydraulic slope, nutrients, and human activities (i.e. dams). Our study revealed that benthic diatoms represented by motile species in ecological guilds are typical consequences of environmental selection in a lotic-

oligotrophic river; this work enables better understanding of the specific contributions of benthic diatoms in biogeochemical cycles in world's large river ecosystems.

Methods

Sample collection

The Yangtze River is the longest river situated wholly in Asia and the third longest in the world, with a drainage basin of 1.8 million km^2. The river is over 6300 km long, has its source in the Qinghai-Tibet Plateau, and flows eastwards into the East China Sea near Shanghai. Over its length, the Yangtze River experiences great changes in landform type and hydrological regime, and supports more than 588 million people [68]. To investigate the seasonal and spatial distributions of diatom communities along the whole river, water and sediment samples were synchronously (i.e. within one week) collected for planktonic and benthic diatom identification at 50 national monitoring stations along the mainstream and six major tributaries of the Yangtze in March (spring) and October (autumn) 2014. In July 2017, we collected water and sediment samples at 12 sites in the river source. In total, 96 water samples and 183 sediment samples were obtained. Except for a very few samples missed due to restrictions of steep terrain and rapid flow as described in a previous study [32], up to four parallel samples were collected in most cases. Further details of the sampling sites are listed in Table S1. At each sampling site, 10L of well-mixed water was collected and then immediately filtered onto 0.22 μm polycarbonate membranes (Millipore, USA) within 24 h. Filtered membranes and sediment samples were stored

461 in the laboratory at -80°C until further analysis took place.

462

463 **DNA extraction, PCR amplification and sequencing**

464 DNA was extracted in triplicate using the FastDNA® SPIN Kit for Soil (MP
465 Biomedicals, USA) following the manufacturer's instructions. The triplicate DNA
466 extracts were mixed together for later PCR amplification. Amplification of the V4
467 region of the 18S rDNA was performed by polymerase chain reaction (PCR) (initial
468 denaturation at 94 °C for 2 min, then 32 cycles of denaturation at 94 °C for 45 s,
469 annealing at 50°C for 45 s, elongation at 72 °C for 60 s, and final elongation at 72 °C
470 for 10 min, 10°C until halted by user) using DIV4for (5'-
471 GCGGTAATTCCAGCTCCAATAG-3') and DIV4rev3 primers (5'-
472 CTCTGACAATGGAATACGAATA-3') [12]. PCR mixtures (20 μL volume) prepared
473 in triplicate contained 2 μl of 10×Buffer, 2 μl of 2.5 mM dNTPs, 0.8 μl of each
474 primer (5 μM), 0.2 μl of rTaq Polymerase, 0.2 μl of BSA, and 1 μl of 10 ng DNA
475 sample. Amplicons were purified using the AxyPrep DNA Gel Extraction Kit
476 (Axygen Bioscience, Union City, CA, U.S.) according to the manufacturer's
477 instructions and quantified using QuantiFluorTM -ST (Promega, U.S.). Afterwards,
478 purified amplicons were pooled in equimolar amounts and sequenced on Illumina
479 MiSeq 2 × 250 PE platform (Majorbio Company, Shanghai, China).

480 Three negative control samples were used to monitor any contamination during the
481 molecular workflow, negative filtration, DNA extraction, and PCR controls, however,

no quantifiable DNA was detected for further analysis.

Bioinformatics analysis

Sequences of diatom 18S rDNA were quality-filtered using QIIME [69] as follows: (i) minimum sequence length of 300 bp, and minimum threshold quality score of Q20; (ii) maximum mismatches of 2 for matching the primer; any reads with ambiguous bases were removed; and (iii) merged pair-ended sequences that overlapped longer than 10 bp into a single sequence. UCHIME was used to remove chimeric sequences and UPARSE was used to cluster Operational Taxonomic Units (OTUs) with 97% similarity cutoff [70].

We built a reference database of 18S rRNA reads composed of 4573 unique diatom sequences. First, we extracted all diatom sequences of 18S rRNA reads from Genbank (http://www.ncbi.nlm.nih.gov/). Second, short reads (less than 100 nucleotides) were refused access to the reference database, and redundant reads were eliminated by cd-hit to increase the taxonomy identification accuracy. Third, sequence alignment was performed by Mafft (ver 7.310) [71], then the sequences were analyzed to construct an approximately-maximum-likelihood phylogenetic tree using FastTree (ver 2.1.10) [72], and any incorrect reads discarded. Finally, a total of 4573 unique sequences were retained in our reference database.

To identify taxonomically the OTUs obtained in this study against known diatom species, the BLASTN [73] program was applied to align clean 18S rRNA reads to the

503 corrected diatom database. Those OTUs with the best BLAST hit scores, not only an
504 e-value $\leq 10^{-5}$ but also identity $\geq 80\%$ with respect to the reference sequence were
505 firstly selected. Then the selected OTUs were checked by means of the phylogenetic
506 tree, and only OTUs with correct taxonomical assignment were retained for further
507 analysis. Clean reads were further assigned to known diatom species based on our
508 reference database.
509 To estimate the community structure for each site, the Mothur program [74] was
510 used to normalize all data sets with respect to the least-well-represented data set
511 (11049 sequences). Alpha diversity indices (chao1, Shannon and Goods coverage)
512 were calculated using QIIME.
513
514 **Statistical analysis**
515 Diatom species that characterize each sample group were identified with Indicator
516 Species Analysis using labdsv and indval packages in R software [75]. Indicator
517 values were calculated based on the relative frequency and relative average
518 abundance of a given species in six types of environmental samples. Species with
519 indicator value $\geqslant$ 0.3 and p-value $\leqslant$ 0.01 were defined as indicator species at
520 Class, Order, Family, and Genus levels. Nonmetric multidimensional scaling (NMDS)
521 was performed to visualize the dissimilarity of different samples based on Bray–
522 Curtis similarity matrices. Analysis of Similarity (ANOSIM) was conducted to test the
523 significance of differences among a priori sampling groups based on environmental

524 parameters. NMDS and ANOSIM statistics were carried out using the vegan package
525 in R. The linear discriminant analysis effect size (LEfSe) [76] was used to discover
526 high-dimensional biomarker and explain taxa difference at different environment
527 conditions. One-way analysis of variance (one-way ANOVA) was carried out to test
528 significance of group differences.
529 Distance-decay patterns of diatom community similarity were described by
530 considering geographical distance from the site location to river mouth among sample
531 sites. Mantel tests were used to examine the Spearman's rank correlation between
532 geographical distance and diatom community similarity using Bray-Curtis distance
533 matrices with 999 permutations in R. The distance of each sampling site was
534 calculated using ArcGIS V10.3 software. The rate of distance-decay of diatom
535 communities was calculated as the slope of ordinary least-squares regression line
536 fitted to the relationship between geographic distance and community similarity.
537 A set of spatial variables was generated through the use of principal coordinates of
538 neighbor matrices (PCNM) analysis based on the longitude and latitude coordinates of
539 each sampling site [77]. The function 'envfit' was run with 999 permutations to select
540 significant variables ($P < 0.05$). Significance testing was then assessed using the
541 'permutest' function based on 999 permutations in R. Partial canonical
542 correspondence analysis (pCCA) was performed to decompose the total variation in
543 diatom community into a pure environmental component, a pure spatial component, a
544 spatially structured environmental component, and residual variation.

545

546 **Ecological guilds classification**

547 Based on their ecological characteristics, diatom species are classified into four
548 ecological guilds (low profile, high profile, motile and planktic guilds) [29–31], which
549 are expected to respond in different ways to nutrient conditions and physical
550 disturbances. A low-profile guild is defined as having high reproduction rate, low
551 nutrient and light availability, and slow-moving diatoms. A high-profile guild
552 possesses characteristics of high resource availability and low disturbance. A motile
553 ecological guild has the ability to move fast and choose the best microhabitat in a
554 given circumstance. A planktic guild adapts to lentic environments and resists
555 sedimentation. We extended these guilds by adding supplemented classifications used
556 in other studies [78,79].

557

558 **Photosynthetically Active Radiation (PAR) divisions**

559 Solar radiation is the most important source of energy required for plant growth. Solar
560 radiation with wavelengths (400-700 nm), called Photosynthetically Active Radiation
561 (PAR) is able to convert light energy into biomass [33] through photosynthesis by
562 plants and algae. Monteith reported the linear correlation between net primary
563 production (NPP) and PAR absorbed by green foliage [80]. Xudong et al. [34] also
564 suggested that the spatial distribution of annually-averaged PAR is complex and
565 inhomogeneous across China, using data for the period 1961−2007. Thus, we define

four zones of PAR intensity in different regions across the Yangtze River basin as follows:

I. Very high, PAR > 32 mol m^{-2} d^{-1};

II. High, 26 < PAR ≤ 32 mol m^{-2} d^{-1};

III. Medium, 23 < PAR ≤ 26 mol m^{-2} d^{-1};

IV. Low, PAR ≤ 23 mol m^{-2} d^{-1}.

Xudong et al. found slowly changing rates of spring-averaged and autumn-averaged PAR occurred in the Yangtze River basin.

Additional files

Additional file 1:

Figure S1. Rarefaction curves of diatom richness per sample at cutoff levels of 3%.

Figure S2. Phylogenetic distribution of reference sequence and abundance of OTUs. The color range displays class level taxonomy information on the node branch. The outmost heatmap indicates relative abundance of OTUs in six sample types: WS, water-spring; SS, sediment-spring; WA, water-autumn; SA, sediment-autumn; WP, water-plateau; SP, sediment-plateau.

Figure S3. Spearman relationships for relative abundance of diatoms and dissolved CO_2 (pCO_2, ppm). Strong correlations between dissolved CO_2 and the planktonic and benthic diatoms are marked in red..

Figure S4. Alpha diversity index per diatom community obtained for each type of

586 sample.

587 **Figure S5.**Nonmetric multidimensional scaling (NMDS) diagram of compositional

588 (Bray-Curtis) dissimilarities between diatom communities for all samples in the

589 Yangtze River..

590 **Figure S6.** ANOSIM statistics concerning differences in diatom communities within

591 and between sample types.

592 **Figure S7.** Diatom genera exhibiting significant seasonal differences in water (**a**) and

593 sediment samples (**b**).

594 **Figure S8.** Proportion of indicator diatoms in each sample type obtained using

595 indicator taxa analysis at Class (**a**), Order (**b**), Family (**c**), and Genus (**d**) levels.

596 **Figure S9.** Biogeographical distribution of diatom species at class level throughout

597 the mainstream of the Yangtze River for: (**a**) water-spring, (**b**) water-autumn, (**c**)

598 sediment-spring, and (**d**) sediment-autumn samples. For comparison, water-plateau

599 samples are displayed in (**a**) and (**b**); sediment-plateau samples are displayed in (**c**)

600 and (**d**).

601 **Figure S10.** Biogeographical distribution of ecological guilds at species level

602 throughout the mainstream of the Yangtze River for all samples. For comparison,

603 water-plateau samples are displayed in (**a**) and (**b**)**;** sediment-plateau samples are

604 displayed in (**c**) and (**d**).

605 **Figure S11.** Relationships between diatom community (Bray–Curtis) similarity and

606 geographic distance for water (**a**) and sediment (**b**) samples. Values of Mantel

607 Spearman correlation (r) and correlation significance (P) are provided. Red lines
608 indicate the ordinary least squares linear regression across all samples.
609 **Figure S12.** Variation in community composition explained by environmental, spatial,
610 and spatially structured environmental component.
611 **Figure S13.** LEfSe cladogram of planktonic (**a**) and benthic (**b**) diatom communities
612 for the three channel slope regions. Diatom taxa with a mean relative abundance of
613 $\geqslant$ 0.1% in all samples, assigned to kingdom (innermost), phylum, class, order,
614 family, and genus (outermost), are used to determine taxa or clades most likely to
615 explain differences between channel slope regions. Differentially abundant taxa
616 (biomarkers) are colored according to the slope regions in which they are most
617 abundant; i.e. red, green and blue circles stand for biomarkers in steep, moderate and
618 mild slope regions.
619 **Figure S14.** Relationships between Shannon diversity and TN:TP for water-spring
620 (**a**), water-autumn (**b**), sediment-spring (**c**), and sediment-autumn (**d**) samples.
621 Annually-averaged TN:TP data during 2005−2014 are used for water samples,
622 whereas monitored data in spring and autumn 2014 are used for sediment samples.
623 Distance relationship of TN:TP for sampling sites along the mainstream is shown in
624 (**e**).
625 **Figure S15.** Significant differences in abundance of benthic diatoms upstream and
626 downstream of Xiluodu Dam (**a**) and Three Gorges Dam (**b**).
627 **Table S1**. Numbers of Indicator species and Top Indicator species across sample sites.

628 **Table S2**. Partial Mantel test for Spearman correlations between community similarity
629 and geographic and environmental distances.
630 **Table S3**. Effects of selected environmental and spatial factors on diatom community
631 composition.
632

633 **Abbreviations**

634 OTUs: operational taxonomic units; HTS: high-throughput sequencing; NMDS:
635 Nonmetric multidimensional scaling; ANOSIM: Analysis of Similarity; one-way
636 ANOVA: One-way Analysis of Variance; LefSe: linear discriminant analysis effect
637 size; PAR: Photosynthetically Active Radiation; TN:TP: ratio of total nitrogen to total
638 phosphate

639 Declarations

640 **Ethics approval and consent to participate**

641 Not applicable.

642 **Consent for publication**

643 Not applicable

644 **Availability of data and material**

645 Complete datasets supporting the findings of this article are in the NCBI Sequence
646 Read Archive (SRA) database (Accession Number: SRP153344).

Competing interests

The authors declare no competing financial interests.

Funding

Financial support from National Natural Science Foundation of China (No. 91647211 and 51721006) is appreciated.

Author's contributions

J.R.N. designed the research. J.W.W. and Q.X.L performed research. J.W.W., Q.X.L., J.R.N. and A.G.L.B. wrote the paper. X.Z. and Y.L. contributed new ideas and information. All of the authors contributed to interpretation of the findings.

Acknowledgments

Support from BIC-ESAT and Majorbio Company (Shanghai, China) is gratefully acknowledged.

References

1. Armbrust EV. The life of diatoms in the world's oceans. Nature. 2009;459:185–92.
2. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science. 1998/07/10. 1998;281:237–40.
3. Sims PA, Mann DG, Medlin LK. Evolution of the diatoms: insights from

666 fossil, biological and molecular data. Phycologia. 2006;45:361–402.
667 4. Mann DG, Droop SJM. Biodiversity, biogeography and conservation of
668 diatoms. Hydrobiologia. 1996;336:19–32.
669 5. Oeding S, Taffs KH. Developing a regional diatom index for assessment and
670 monitoring of freshwater streams in sub-tropical Australia. Ecol. Indic. 2017;80:135–
671 46.
672 6. Cermeño P, Falkowski PG. Controls on Diatom Biogeography in the ocean.
673 Science. 2009;325:1539–41.
674 7. Orlando Necchi JR. River Algae. Necchi JR O, editor. Cham: Springer
675 International Publishing; 2016.
676 8. Kelly MG. Use of the trophic diatom index to monitor eutrophication in
677 rivers. Water Res. 1998;32:236–42.
678 9. Zimmermann J, Glöckner G, Jahn R, Enke N, Gemeinholzer B.
679 Metabarcoding vs. morphological identification to assess diatom diversity in
680 environmental studies. Mol. Ecol. Resour. 2015;15:526–42.
681 10. Evans KM, Wortley AH, Mann DG. An Assessment of Potential Diatom
682 “Barcode” Genes (cox1, rbcL, 18S and ITS rDNA) and their Effectiveness in
683 Determining Relationships in Sellaphora (Bacillariophyta). Protist. 2007;158:349–64.
684 11. Moniz MBJ, Kaczmarska I. Barcoding of Diatoms: Nuclear Encoded ITS
685 Revisited. Protist. 2010;161:7–34.
686 12. Visco JA, Apotheloz-Perret-Gentil L, Cordonier A, Esling P, Pillet L,

687 Pawlowski J. Environmental Monitoring: Inferring the Diatom Index from Next-
688 Generation Sequencing Data. Environ. Sci. Technol. 2015;49:7597–605.
689 13. Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, et al.
690 Insights into global diatom distribution and diversity in the world's ocean. Proc. Natl.
691 Acad. Sci. U. S. A. 2016;113:E1516–25.
692 14. Boyd PW, Watson AJ, Law CS, Abraham ER, Trull T, Murdoch R, et al. A
693 mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron
694 fertilization. Nature. 2000;407:695–702.
695 15. Tsuda A, Takeda S, Saito H, Nishioka J, Nojiri Y, Kudo I, et al. A
696 mesoscale iron enrichment in the western subarctic pacific induces a large centric
697 diatom bloom. Science. 2003;300:958–61.
698 16. Heinsalu A, Alliksaar T, Leeben A, Nõges T. Sediment diatom
699 assemblages and composition of pore-water dissolved organic matter reflect recent
700 eutrophication history of Lake Peipsi (Estonia/Russia). Hydrobiologia. 2007;584:133–
701 43.
702 17. Kitner M, Poulíčková A. Littoral diatoms as indicators for the
703 eutrophication of shallow lakes. Hydrobiologia. 2003;506–509:519–24.
704 18. Liu J, Soininen J, Han B-P, Declerck SAJ. Effects of connectivity,
705 dispersal directionality and functional traits on the metacommunity structure of river
706 benthic diatoms. J. Biogeogr. 2013;40:2238–48.
707 19. Kheiri S, Solak CN, Edlund MB, Spaulding S, Nejadsattari T, Asri Y, et al.

708 Biodiversity of diatoms in the Karaj River in the Central Alborz, Iran. Diatom Res.
709 2018;33:355–80.
710 20. Centis B, Tolotti M, Salmaso N. Structure of the diatom community of the
711 river Adige (North-Eastern Italy) along a hydrological gradient. Hydrobiologia.
712 2010;639:37–42.
713 21. Rimet F. Benthic diatom assemblages and their correspondence with
714 ecoregional classifications: Case study of rivers in North-Eastern France.
715 Hydrobiologia. 2009;636:137–51.
716 22. Dalu T, Wasserman RJ, Magoro ML, Mwedzi T, Froneman PW, Weyl
717 OLF. Variation partitioning of benthic diatom community matrices: Effects of
718 multiple variables on benthic diatom communities in an Austral temperate river
719 system. Sci. Total Environ. 2017;601–602:73–82.
720 23. Kireta AR, Reavie ED, Sgro G V, Angradi TR, Bolgrien DW, Hill BH, et
721 al. Planktonic and periphytic diatoms as indicators of stress on great rivers of the
722 United States: Testing water quality and disturbance models. Ecol. Indic.
723 2012;13:222–31.
724 24. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green
725 JL, et al. Microbial biogeography: putting microorganisms on the map. Nat. Rev.
726 Microbiol. 2006;4:102–12.
727 25. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond
728 biogeographic patterns: processes shaping the microbial landscape. Nat. Rev.

729 Microbiol. 2012;10:1–10.

730 26. Keck F, Franc A, Kahlert M. Disentangling the processes driving the
731 biogeography of freshwater diatoms: A multiscale approach. J. Biogeogr.
732 2018;45:1582–92.

733 27. Heino J, Bini LM, Karjalainen SM, Mykrä H, Soininen J, Vieira LCG, et
734 al. Geographical patterns of micro-organismal community structure: Are diatoms
735 ubiquitously distributed across boreal streams? Oikos. 2010;119:129–37.

736 28. Soininen J. Determinants of benthic diatom community structure in boreal
737 streams: The role of environmental and spatial factors at different scales. Int. Rev.
738 Hydrobiol. 2004;89:139–50.

739 29. Passy SI. Diatom ecological guilds display distinct and predictable
740 behavior along nutrient and disturbance gradients in running waters. Aquat. Bot.
741 2007;86:171–8.

742 30. Rimet F, Bouchez A. Life-forms, cell-sizes and ecological guilds of
743 diatoms in European rivers. Knowl. Manag. Aquat. Ecosyst. 2012;01.

744 31. B-Béres V, Lukács Á, Török P, Kókai Z, Novák Z, T-Krasznai E, et al.
745 Combined eco-morphological functional groups are reliable indicators of colonisation
746 processes of benthic diatom assemblages in a lowland stream. Ecol. Indic.
747 2016;64:31–8.

748 32. Liu T, Zhang AN, Wang J, Liu S, Jiang X, Dang C, et al. Integrated
749 biogeography of planktonic and sedimentary bacterial communities in the Yangtze

River. Microbiome. 2018;6:1–14.

33. Jacovides CP, Timvios FS, Papaioannou G, Asimakopoulos DN, Theofilou CM. Ratio of PAR to broadband solar radiation measured in Cyprus. Agric. For. Meteorol. 2004;121:135–40.

34. Zhu X, He H, Liu M, Yu G, Sun X, Gao Y. Spatio-temporal variation of photosynthetically active radiation in China in recent 50 years. J. Geogr. Sci. 2010;20:803–17.

35. Montagnes DJS, Franklin DJ. Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: Reconsidering some paradigms. Limnol Ocean. 2001;46:2008–18.

36. Yu G-A an, Brierley G, Huang HQ, Wang Z, Blue B, Ma Y. An environmental gradient of vegetative controls upon channel planform in the source region of the Yangtze and Yellow Rivers. Catena. 2014;119:143–53.

37. Chen Z, Li J, Shen H, Zhanghua W. Yangtze River of China: Historical analysis of discharge variability and sediment flux. Geomorphology. 2001;41:77–91.

38. Redfield AC, Ketchum BH, Richards FA. The influence of organisms on the composition of seawater. Sea. New York, USA: Wiley-Interscience; 1963.

39. Assmy P, Smetacek V, Montresor M, Klaas C, Henjes J, Strass VH, et al. Thick-shelled, grazer-protected diatoms decouple ocean carbon and silicon cycles in the iron-limited Antarctic Circumpolar Current. Proc. Natl. Acad. Sci. U. S. A. 2013;110:20633–8.

40. Leblanc K, Quéguiner B, Diaz F, Cornet V, Michel-Rodriguez M, Durrieu de Madron X, et al. Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. Nat. Commun. 2018;9:953.

41. Tréguer P, Bowler C, Moriceau B, Dutkiewicz S, Gehlen M, Aumont O, et al. Influence of diatom diversity on the ocean biological carbon pump. Nat. Geosci. 2018;11:27–37.

42. Smetacek V, Ban S, Burns G, Castel J, Chaudron Y, Christou R, et al. Diatoms and the Ocean Carbon Cycle. Protist. 1999;150:25–32.

43. Duleba M, Ector L, Horváth Z, Kiss KT, Molnár LF, Pohner Z, et al. Biogeography and Phylogenetic Position of a Warm-stenotherm Centric Diatom, Skeletonema potamos (C.I. Weber) Hasle and its Long-term Dynamics in the River Danube. Protist. 2014;165:715–29.

44. Sims PA, Crawford RM. Earliest records of Ellerbeckia and Paralia from Cretaceous deposits: a description of three species, two of which are new. Diatom Res. 2017;32:1–9.

45. Zong Y. Implications of paralia sulcata abundance in scottish isolation basins. Diatom Res. 1997;12:125–50.

46. Jianjun Z, Man Z. Effect of dams on the regime of the mid-lower Yangtze River runoff and countermeasures.pdf. J. Lake Sci. 2018;30:1471–88.

47. Souffreau C, Verbruggen H, Wolfe AP, Vanormelingen P, Siver PA, Cox EJ, et al. A time-calibrated multi-gene phylogeny of the diatom genus Pinnularia.

792 Mol. Phylogenet. Evol. 2011;61:866–79.
793 48. Arai H, Fukushima T. Impacts of long-term increase in silicon
794 concentration on diatom blooms in Lake Kasumigaura, Japan. Ann. Limnol.
795 2014;50:335–46.
796 49. Cohn SA, Farrell JF, Munro JD, Ragland RL, Weitzell RE, Wibisono BL.
797 The effect of temperature and mixed species composition on diatom motility and
798 adhesion. Diatom Res. 2003;18:225–43.
799 50. Hauptmann AL, Markussen TN, Stibal M, Olsen NS, Elberling B, Bælum
800 J, et al. Upstream Freshwater and Terrestrial Sources Are Differentially Reflected in
801 the Bacterial Community Structure along a Small Arctic River and Its Estuary. Front.
802 Microbiol. 2016;7:1474.
803 51. Domingues RB, Anselmo TP, Barbosa AB, Sommer U, Galvão HM. Light
804 as a driver of phytoplankton growth and production in the freshwater tidal zone of a
805 turbid estuary. Estuar. Coast. Shelf Sci. 2011;91:526–35.
806 52. Buma AGJ, Engelen AH, Gieskes WWC. Wavelength-dependent induction
807 of thymine dimers and growth rate reduction in the marine diatom Cyclotella sp.
808 exposed to ultraviolet radiation. Mar. Ecol. Prog. Ser. 1997;153:91–7.
809 53. Depauw FA, Rogato A, D'Alcalá MR, Falciatore A. Exploring the
810 molecular basis of responses to light in marine diatoms. J. Exp. Bot. 2012;63:1575–
811 91.
812 54. Bopp L, Aumont O, Cadule P, Alvain S, Gehlen M. Response of diatoms

813 distribution to global warming and potential implications : A global model study.
814 Geophys. Res. Lett. 2005;32:1–4.
815 55. Peters F, Arin L, Marrasé C, Berdalet E, Sala MM. Effects of small-scale
816 turbulence on the growth of two diatoms of different size in a phosphorus-limited
817 medium. J. Mar. Syst. 2006;61:134–48.
818 56. Mitrovic SM, Chessman BC, Davie A, Avery EL, Ryan N. Development of
819 blooms of Cyclotella meneghiniana and Nitzschia spp. (Bacillariophyceae) in a
820 shallow river and estimation of effective suppression flows. Hydrobiologia.
821 2008;596:173–85.
822 57. Han L, Xu B, Qi F, Chen Z. Effect of nitrogen/phosphorus concentration
823 on algal organic matter generation of the diatom Nitzschia palea: Total indicators and
824 spectroscopic characterization. J. Environ. Sci. (China). 2016;47:130–42.
825 58. Tantanasarit C, Englande AJ, Babel S. Nitrogen, phosphorus and silicon
826 uptake kinetics by marine diatom Chaetoceros calcitrans under high nutrient
827 concentrations. J. Exp. Mar. Bio. Ecol. 2013;446:67–75.
828 59. Zeng Q, Qin L, Bao L, Li Y, Li X. Critical nutrient thresholds needed to
829 control eutrophication and synergistic interactions between phosphorus and different
830 nitrogen sources. Environ. Sci. Pollut. Res. 2016;23:21008–19.
831 60. Cotner, Jr JB, Wetzel RG. Uptake of dissolved inorganic and organic
832 phosphorus compounds by phytoplankton and bacterioplankton. Limnol. Oceanogr.
833 1992;37:232–43.

834 61. Kunz MJ, Wüest A, Wehrli B, Landert J, Senn DB. Impact of a large
835 tropical reservoir on riverine transport of sediment, carbon, and nutrients to
836 downstream wetlands. Water Resour. Res. 2011;47:W12531.
837 62. Graf WL. Downstream hydrologic and geomorphic effects of large dams
838 on American rivers. Geomorphology. 2006;79:336–60.
839 63. Tekwani N, Majdi N, Mialet B, Tornés E, Urrea-Clos G, Buffan-Dubau E,
840 et al. Contribution of epilithic diatoms to benthic-pelagic coupling in a temperate
841 river. Aquat. Microb. Ecol. 2013;69:47–57.
842 64. Jäger CG, Diehl S. Resource competition across habitat boundaries:
843 asymmetric interactions between benthic and pelagic producers. Ecol. Monogr.
844 2014;84:287–302.
845 65. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The
846 River Continuum Concept. Can. J. Fish. Aquat. Sci. 1980;37:130–7.
847 66. Ward JV, Stanford JA. The serial discontinuity concept of lotic ecoystems.
848 In: Fontaine, T.D., Bartell SM, editor. Dynamics of Lotic Ecosystems. Ann Arbor
849 Science Publications. Ann Arbor, Michigan; 1983.
850 67. Jäger CG, Borchardt D. Longitudinal patterns and response lengths of algae
851 in riverine ecosystems: A model analysis emphasising benthic-pelagic interactions. J.
852 Theor. Biol. 2018;442:66–78.
853 68. Qu S, Wang L, Lin A, Zhu H, Yuan M. What drives the vegetation
854 restoration in Yangtze River basin, China: Climate change or anthropogenic factors?

855 Ecol. Indic. 2018;90:438–50.

856 69. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD,
857 Costello EK, et al. QIIME allows analysis of high-throughput community sequencing
858 data. Nat. Methods. 2010;7:335–6.

859 70. Edgar RC. UPARSE: highly accurate OTU sequences from microbial
860 amplicon reads. Nat. Methods. 2013;10:996–8.

861 71. Yamada KD, Tomii K, Katoh K. Application of the MAFFT sequence
862 alignment program to large data - Reexamination of the usefulness of chained guide
863 trees. Bioinformatics. 2016;32:3246–51.

864 72. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-
865 likelihood trees for large alignments. PLoS One. 2010;5:e9490.

866 73. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al.
867 Gapped BLAST and PSI-BLAST: A new generation of protein database search
868 programs. Nucleic Acids Res. 1997;25:3389–402.

869 74. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et
870 al. Introducing MOTHUR: open-source, platform-independent, community-supported
871 software for describing and comparing microbial communities. Appl Env. Microbiol.
872 2009;75:7537–41.

873 75. Dufrtne M, Legendre P. Species assemblages and indicator species: The
874 need for a flexible asymmetrical approach. Ecol. Monogr. 1997;67:345–66.

875 76. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al.

876 Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60.
877 77. Borcard D, Legendre P. All-scale spatial analysis of ecological data by
878 means of principal coordinates of neighbor matrices. Ecol Model. 2002;153:51–68.
879 78. Berthon V, Bouchez A, Rimet F. Using diatom life-forms and ecological
880 guilds to assess organic pollution and trophic level in rivers: A case study of rivers in
881 south-eastern France. Hydrobiologia. 2011;673:259–71.
882 79. B-Béres V, Török P, Kókai Z, Lukács Á, T-Krasznai E, Tóthmérész B, et
883 al. Ecological background of diatom functional groups: Comparability of
884 classification systems. Ecol. Indic. 2017;82:183–8.
885 80. Monteith JL. Solar-Radiation and Productivity in Tropical Ecosystems. J.
886 Appl. Ecol. 1972. p. 747–66.
887

Figure legends

Figure 1. Map of sampling sites in the Yangtze River covering the actual sinuous channel reach of length 6030 km (equivalent to 1.83 times the 3290 km straight line from start to end sampling sites). The molecular method provides insight into the biogeographic pattern of diatoms along the mainstream of the Yangtze River. Photosynthetically Active Radiation (PAR), temperature, channel slope, and nutrients characterized by ratio of total nitrogen to total phosphate (TN:TP) are used to interpret the biogeographic patterns of diatom communities.

Figure 2. (**a**) Circular visualization of dominant diatoms at genus level in six sample types. Inner circular diagram shows relative abundance of different diatom genus in six sample types. Only the dominant genus with a mean relative abundance of $\geqslant$ 1% in all samples is depicted. The width of ribbons for each diatoms is directly proportional to their relative abundance in each sample type. Similarly, different colored ribbons of different width for each sample type describe the distribution of different genera. (**b**) Representative diatoms genera in different landform types from the river source to mouth along the Yangtze River.

Figure 3. LEfSe cladogram of planktonic (**a**) and benthic (**b**) diatom communities from four PAR regions. Diatom taxa with a mean relative abundance of $\geqslant$ 0.1% in all samples, assigned to kingdom (innermost), phylum, class, order, family, and genus (outermost), are used to determine taxa or clades most likely to explain differences between PAR regions. Differentially abundant taxa (biomarkers) are colored by their

most abundant PAR regions, i.e. red, green, blue and purple circles stand for biomarkers in regions of very high, high, medium and low abundance. Orange and blue circles display the average alpha-diversity (Chao1) of planktonic and benthic diatoms respectively, in different photosynthetically active radiation (PAR) regions, with size corresponding to the Chao1 index (**c**).

Figure 4. Relationships between community similarity and river channel slope for water-spring (**a**), water-autumn (**b**), sediment-spring (**c**)**,** and sediment-autumn (**d**) samples. Values of Mantel Spearman correlation (r) and correlation significance (P) are also provided. Gray lines denote ordinary least squares linear regression fits across all samples. Spatial distributions of ecological guilds for different channel slopes are shown in (**e**).